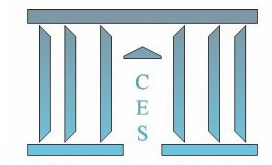
UNIVERSITÉ PARIS 1 – PANTHÉON - SORBONNE –

Documents de Travail du
Centre d'Economie de la Sorbonne

CES Working Papers



Are public transport improvements endogenous with respect to employment and income location in a city?

Carlos Augusto OLARTE BACARES

2014.12

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
http://centredeconomiesorbonne.univ-paris1.fr/
ISSN : 1955-611X

Are public transport improvements endogenous with respect to employment and income location in a city?*

Carlos Augusto Olarte Bacares†

February 1, 2013

Abstract

Previous research has proved the existence of a causal relationship between the concentration of jobs in a city and the income of inhabitants. Other researchers have studied the close and nearly causal relationship between those variables and the infrastructure such as highways in different zones of a city. Nevertheless, no one study has taken into account the degree to which each area of a city benefits from the latest improvements to public transport. The aim of this research is to analyse the relationship between the size of the labour market, the income and the employment concentration with respect to improvements to public transport (Transmilenio) in Bogota . The degree of enhancement of public transport in a zone is suspected to be endogenous. Through the use of OLS estimations and then 2SLS, the validation of endogeneity provides sufficient tools to infer causality of improvement of public transport. The size of companies, defined by the number of jobs they offer, plays the role of instrumental variable. In essence, the number of jobs, the size of the labour market and income are largely defined by the level of improvement to urban public transport in each zone of the city but the causality relationship changes depending on the size of companies established in each zone. In the case of Bogotá, public transport improvements seams to have a causality relationship with the income of inhabitants in each zone and the number of jobs, and this changes with respect to the size of enterprises. In contrast, the size of the labour market, defined as the number of jobs reachable in a specific time, is not determined by the degree of the presence of public transport enhancement.

JEL Codes: J68, R12, R23

Keywords: Causality; Improvements of public transports; Endogeneity; Effective size of labor market; Size of enterprises

*This paper was accepted and presented to the refereed "Young Scientists' Session" at the 53 rd ERSA congress. Palermo August 2013.

†Centre d'Economie de la Sorbonne, Université Paris I, Panthéon-Sorbonne, CNRS, 106 boulevard de l'Hôpital, 75647 Paris Cedex 13, France. E-mail: carlos.olarte-bacares@univ-paris1.fr

1 INTRODUCTION

1.1 A brief theoretical contextualization

Connections between commuting time and localization of jobs in urban areas have been widely studied by several researchers (Cervero, 2000; Hansen, 1978; Prud'Homme, 1999). The number of jobs that can be reached in a specific period of time, which can be directly related to the concept of job accessibility (Kay Axhausen, 2008), represents a relevant topic in urban studies and transportation research. Time of urban travel, distance from homes to jobs and urban structure seem to hoard the attention of the majority of researchers involved in urban and urban transport studies.

Inhabitants choose to live in a specific zone of a city taking into account numerous elements of houses and neighbourhoods according to the theory of hedonic price (Rosen 1974). The area of the house in square metres, the comfort of the house, the materials which with houses are built, seem to be very important to buyers. However, in the past decades, due to the increasing congestion in cities, inhabitants have been giving increasing importance to the amenities of the neighbourhood and the zone where their houses are placed (Glaeser, 1999; Brueckner, 2001; Putman, 2000). The better and more amenities the neighbourhood offers, the more expensive the house will be. In other words, the actual configuration of cities and, thus, inhabitants' choices about the zone in which they decide to live, does not depend just on their desires; more precisely, it depends on the facilities of the zone and hence on the price of the amenities offered in the area.

Therefore, it is not strange to read that disparities of income in a city represent one of the principal causes of the socio-economic gap between different zones in a city (Glaeser and Thysse, 2004; Cao, Morkarian and Handy 2007). Because of their high purchasing power, people and companies with a high level of income used to have more possibilities to choose the zone of the city they want to dwell; they have more flexibility to move and change their place where they live (Bruekner, 2002; Glaeser, Rosenthal and Strange 2010; Anas, Arnott and Small, 1998).

In the 70s and 80s, studies of American cities showed that, because of the widespread usage of the private car, people preferred to live far from all kinds of centres of the city (commercial centre, job centre, government centre, etc.) and inhabited suburbs alongside people with the same level of income and the same level of education. They preferred suburban spaces with less noise from cars, less congestion and better security, which offered big parks and more recreation spaces than zones closer to the centre of the city (Giuliano and Small, 1991; Fujita, 1999). In addition, as noted above, those kinds of suburbs have the most marked homogeneity of social classes. Due to high property prices, people with a low level of income will not be able to live there. This phenomenon was

generally perceived in cities both in developed and developing countries. It can be perceived as a phenomenon of cities in general.

This research is focused on the city of Bogotá, the capital of Colombia. As some studies have shown for other cities of developed countries, the city centre of Bogota had a decline from the urban and planning perspective. A spatial mismatch began to settle down in the city (Bureckner and Zenou, 2003).

1.2 Some characteristics of Bogotá

Before the 1960s, provision of public transport was led by the public administration. Buses, taxis and tramway lines composed the urban transport system of the city, with the tramway as the "spinal column". The city centre was a desirable place to live.

Nevertheless, because of Bogotá's huge public debt and due to various social factors[1], tramway took ten years to close and the public administration decided to privatize the public transport system. Since the 60s and due to the destruction of many buildings during several civil manifestations, the centre of Bogotá stopped being a desirable place to live. The management of public transport was given over to private entrepreneurships which were just focused on the development of more lines of buses and not on the urbanization or planning of the city. Public administration was supposed to have a regulatory role but, because of the lobbying power of private entrepreneurships, that role was not assured by the administration.

Today, the city's public transport system continues to be managed by private entrepreneurships assembled in about 60 private enterprises assuring the provision of the service in the city[2]. The public administration does not participate in utilities. Since the late 1960s, the administration's regulatory role has involved determining the different lines of buses needed in zones of the city; after a call for tender, it assigns those lines to "winning" enterprises. This mode of management has contributed to the expansion of the public transport system and thereby the major spread of the city in the past 40 years.

In the past decades downtown has become older, not safe, noisy during the day and abandoned and dark during the night. Because of these characteristics, during this period, prices of houses at downtown decreased significantly; people with low income opted to live in the city centre. However, this does not mean that they live only in the centre of the city. They also live in suburbs, far from the centre of the city, but whit characteristics widely different than those of zones that "rich" people used to inhabit.

[1] Public debt and "El Bogotazo": civil insurrection after the homicide of a very important politician (Jorge Eliecer Gaitan) a future president of Colombia at 1948.

[2] From Secretaria de Movilidad del Distrito

That type of behaviour encouraged the development of inner cities (Small, 1999; Fujita, 1989; Duranton and Anas, 2001) and also increasing expansion of urban areas. Generally, suburbs where people with a low level of income live do not have good coverage of public services. The water and sewerage services are not provided in the same proportion. Clean water, electricity and even paved roads and, therefore, urban transportation are not offered in the same proportion.

The population continues to grow rapidly in the urban area and, hence, congestion is growing faster than before. Nowadays, people take more time that they are ready to spend from their houses to their jobs. Cities are still to expand its boundaries and "exclusive" suburbs are now too far from the city and, worse, from work. As a consequence of the sprawl and the consistent lack of planning policies for decades, Bogotá has suffered big problems with mobility and its transportation system. Densification of downtown was one of the principal goals of the mayor elected in 2011[3]. The continuous reduction of accessibility, severe congestion on some main roads of the city, long travel times and poor road network conditions became the most relevant problems for the city. These consequences have been called by some researchers the "spatial mismatch" (Kain, 1968, 1993; Ihlandfeldt and Sjoquist, 1991).

At the end of the 1990s, aware of the city's great need, the major of Bogotá decided to improve and innovate public infrastructure of public through the creation of a transport system called "Transmilenio", which has been the most efficient BRT system ever seen in Bogotá (Chaparro and Irma 200;, Hidalgo and Sandoval 2001).

Reorganization of the transport network and restructuring of planning policies are now at the core of the public debate. Today, the public administration participates in the roles of manager and inspector; changes in the social and urban structure have taken place in Bogotá in the past 15 years. Most neighbourhoods (60%) are now connected directly by Transmilenio (TM). Most zones (excluding the suburbs) will be substantially interconnected with downtown areas by TM[4].

Consequently a shift in decisions to live in the suburbs took place. In the past 15 years, people have shown increasing willingness to inhabit the centre of the city because of proximity to their jobs. In Bogotá, more than 30% of jobs are located downtown. This fact had as a consequence a "redensification" phenomenon in the city centre by people with the highest incomes as well as increased prices in the centre. People prefer to live near to their workplace and not an hour and a half from it[5]. Thus, the connections between income, job accessibility, employment and public transport improvement seem to be very tight.

[3] Gustavo Petro was elected for the period 2012-2016

[4] Source: Secretaria de movilidad del Distrito

[5] Time that they normally take if they live in suburbs. Encuesta de movilidad del distrito 2005. Secretaria de movilidad del distrito

Some earlier studies (Immerluck, 1998; Gao, 2006; Gao, Mokhtarian and Johnston, 2008), have demonstrated the dependence and even the causal relationships among accessibility to jobs, employment, income and auto ownership. In this study, I propose to demonstrate the differences in these relationships among zones which have benefited from improvements to the transport network and those that have not had enhancements.

Some studies about hedonic prices and their relationship with public transport reveal that the distance between houses and public transport networks can have positive or negative effects on houses' hedonic prices[6]. Regarding the city of Bogotá, Mendieta and Perdomo proved that "the average elasticity proximity of TM, price of the land are -0.36, -0.55 and -1.13 for the first, second and third stages of TM in the same order"[7]. In parallel, Gutierrez (2011) suggests that amenities of zones, accessibility, economies of scale and other zonal characteristics contribute to explain the configuration of employment centres.

Taking into consideration those studies and following some theoretical and empirical studies (Gao, 2006; Gao, Mokhtarian and Johnston, 2008), this work expects to find that the presence of improvement to public transport in each zone of Bogota has a significant impact on the level of income, the number of jobs and the level of what Prud'Homme and Lee (1999) called "The Effective Size of Labour Market" ($ESLM$) in each zone. Differences of income, employment and $ESLM$ between zones with improvement to their transport networks and zones without the presence of Transmilenio are expected. This research tends to prove that these three variables are not only deeply related to improvements made to public transport, but there also exists an important endogeneity between the level of enhancement of public transport on these three variables. If endogeneity among improvements and other variables is verified, it will let us infer the existence of a causal relation from improvement to income, employment and $ESLM$.

This paper is structured in three main sections, as follows: we first explain the research methodology of the conceptual framework of the subject. We define our instrumental variable (IV) and thus the instruments we use in our three models. We then introduce the core phase of our analysis, which focuses on the demonstration of endogeneity. It may show that there is a causal relation between these factors. In fact, from a theoretical point of view, an endogenous variable is a factor which is almost determined by the importance of other variables in a specific model. Generally, factors can be affected directly by other factors presented in the model but also by factors that could not be included in the system. This means that we can find some degrees of endogeneity (partial or total endogeneity).

[6] In the case of Bogota the distance is measure with respect to the nearest BRT station, Transmilenio.

[7] Mendieta, J. C. y Perdomo, J. A. (2007). "Especificación y estimación de un modelo de precios hedónico espacial para evaluar el impacto de Transmilenio sobre el valor de la propiedad en Bogotá", Documento CEDE, Centro de Estudios sobre Desarrollo Económico, 22, ISSN 1657-5334

As previously stated, our purpose is to find and reaffirm the theoretical endogeneity between improvements to the public transport system and job accessibility, income and employment[8]. Finally, in order to better understand the model's results, a discussion takes place in the fourth section. In this way we will be able to conclude whether a causal relationship between these variables is plausible.

2 CONCEPTUAL FRAMEWORK AND RESEARCH METHODOLOGY

Earlier studies (Thompson, 1997; Sanchez, 1999; Shen and Sanchez, 2005) have shown relationships between employment, income, job accessibility and auto ownership. This research assumes these relationships and aims to demonstrate the importance of the enhancement of public transport provided by the construction of TM on the level of income, number of jobs and degree of accessibility to jobs defined by the ESLM.

According to Kawabata (2003), interdependence between these variables can be expected. In effect, the improvement of urban transport directly influences the index of job accessibility. On the other hand, an upgrading of job accessibility will boost the average income because of the gain in time and the reduction of commuting costs, which will also have a direct effect on the number of jobs. Indeed, the reduction in commuting costs and travel times will bring inhabitants closer from their employments (Prud'homme and Lee, 1999)[9]. In our analysis we assume that this proximity and the reduction of commuting time is due to the presence of more Transmilenio stations.

2.1 How to mediate with endogeneity

To demonstrate whether there is a causal relation, a deep study about the endogeneity of those variables must be done. We will first make an OLS model for each variable with respect to the other control variables, in order to find the degree of relation between those variables when they are dependent or independent factors. This will begin with an examination study of correlations and the variance inflation factors for each variable on each model to determine if the mode could have some multicollinearity problems.

After doing those analyses, the research will proceed with the measurement of possible causal effects. When causality is evoked, the three possible reasons proposed by Kenny (1979) have to be considered:

[8] In the model employment will be taken as the total of jobs on each zone of the city.
[9] See the analysis about the relation between the Speed, the Sprawl and the Spread.

- Factor x must temporally precede the dependent variable (not always)
- Factor x is correlated with the dependent variable
- The relation between the independent and the dependent variable must not be explained by other causes.

Condition 1 can be necessary but is not a sufficient condition to establish causality. Even so, due to our type of data, we will not demonstrate this condition because we are not doing a time series analysis. Our analysis is statistical, hence factor x will not temporally precede the dependent variable. Condition 2 needs, from a statistical point of view, a relationship between the explanatory variable and dependent variable (which is supposed to have been shown before). The third condition supposes the independent variable and dependent variable have a relation that depends on other reasons. Then the independent variable is endogenous and the coefficient of this variable does not reflect a simple correlation. In addition, the dependent variable could also explain the independent variable (reverse causality). To summarize, endogeneity of variables means that there is a presence of functions explanatory variables in the system.

According to Antonakis et al. (2010), independent variables might be endogenous for many reasons. Endogeneity may be caused by the omission of some variables that explain an independent variable. Finally, the fact that independent and dependent variables may be collected from the same rating source could also be a reason of endogeneity among other reason we will not discuss in this issue.

From a mathematical point of view, one or some explanatory variables of a model may be endogenous because they can be correlated with the error term of the system (condition 3). If this is the case, the least square estimator found is not consistent. To have a consistent estimator, researchers have to find valid instruments of the variable suspected to be endogenous. Subsequently, using the appropriate variables, a two-stage least squares (2SLS) model will give a better, consistent estimator that will reflect the "real" influence of the endogenous variable on the explained variable.

After doing the 2SLS analysis, the analysis will continue with a test and correction of the potential endogeneity. Then it is supposed that the results will expose enough tools to conclude if there is endogeneity between variables and, thus, if a causal relationship is verified or not.

2.2 Models and variables structure

As was noted before, it is assumed that there is a relationship between the enhancement of public transport provided by the construction of TM, the level of income, the number of jobs and the degree of accessibility of jobs. Note that there are three dependent variables (three different models to test) and one variable suspected to be endogenous.

2.2.1 Endogenous variable and choice of instrumental variables (IV)

Improvement of public transport ($Improvement_i$) will be the "potential" endogenous variable and will be denoted by the number of Transmilenio stations in each zone; ($Improvement_i$) = 0 means that there is no Transmilenio station in the zone.

The choice of the instruments was made after a simple GIS analysis of the configuration of dependent variables and the instruments. Annexes 6 and 7 show that even if the workforce has a "suspected" degree of relationship with employment and income (figure 1 and annex 4), their scattering across the city seems to be randomized and thus, there is not a big correlation between those variables. In effect, it can be seen that there are zones where the employment is condensed but not for working force and vice versa. The same case is perceived concerning the size of enterprises and independent variables. The concentration of enterprises and their size seems to be random and there is not a direct relation with income, the $ESLM$ or the number of jobs. Analogically, these remarks could join the conclusions of Wenglenski (2006) with respect to the concentration of jobs and the degree of accessibility equality for people of low income inasmuch as there is not a clear relationship regarding the concentration of these elements in the city.

2.2.2 The model: three studied scenarios

This research tries to adapt the model presented in some precedent studies (Cao, Mokhatarian and Handy, 2007; Cervero, 2003) with available data collected for the city of Bogotá. It is assumed that the presence of a Transmilenio station in each zone depends on the workforce in each zone i, WF_i, and on the number of enterprises in each zone, $Enter_{si}$, with respect to their size s. Variables such as employment, income, area and others that could be taken into consideration are supposed to be exogenous in order to avoid multicollinearity problems in main models. Remember that IV tries to capture the effects of exogenous variables that are not taken into consideration in the "main" models, and so on dependent variables. But what were the reasons to choose instrumental variables?

Regarding the number of enterprises, $Enter_{si}$, because of the data available, I decided to classify them into three different groups according to the number of employees and also taking into consideration some details of methodology from the Colombian government[10]. Enterprises with less than ten employees are named "microenterprises" (MiE). Enterprises with more than ten employees and less than 50 employees are called small and middle-sized enterprises (SME). Finally, enterprises with more than 50 employees are considered as large enterprises (LE). The reason to differentiate companies from each other

[10] DANE, Departamento Administrativo Nacional de Estadística in spanish

regarding size is to take into consideration the number of enterprises in each zone, but also the potential jobs that those companies offer. For each model, three different results will be proposed with respect to the number of companies, which are classified into three groups in accordance with the number of employees they have.

Income, employment and job accessibility are the dependent variables to regress.

In this study, job accessibility will be denoted by the Effective Size of Labour Market, $ESLM$, which is the number of jobs reachable for people commuting from one zone to another in a specific time frame. The $ESLM$ was already computed in previous research[11] following the methodology of Prud'homme and Lee (1999). Results are given on three different intervals of time (40 minutes, 50 minutes and 60 minutes)[12], so I test three different levels of $ESLM$ and hence three different models, one per interval of time. A deeper description of this variable is formulated in section 3.

The variable suspected to be endogenous is defined as follows:

$$impro = \pi_0 + \pi_1 Dens_i + \pi_2 WF_i + \pi_3 Enter_{si} + u \tag{1}$$

Models and relationships I propose are:

Model 1 :

$$\begin{aligned} ESLM_{ti} &= \beta_0 Cons + \beta_1 impro_i + \beta_2 Y_i + \beta_3 Pop_i + \beta_4 Area_i + \beta_5 Dist_i \\ &\quad + \beta_6 Car_i + \beta_7 X_i + e \end{aligned} \tag{2}$$

where $Cons$, is the constant variable of the model;

$ESLM_{ti}$, designates the number of reachable jobs for people commuting from zone i in a specific time t. It is the "proxy" of job accessibility;

$Impro_i$, denotes the number of Transmilenio stations in each zone of the city. The more stations the zone has, the better will be the improvement of public transport in these zones; no presence of stations in the zone means that there is no improvement or no presence of TM in i;

[11] Olarte, C. (2012) "Heterogeneity of social classes and job accessibility: implications of transport policies in Bogota" Working paper, Centre d'Economie de la Sorbonne (CES).

[12] I suppose that when people take less than 30 minutes to reach their jobs, they prefer do it by walk and not by public transports. In fact, as it is shown on annex 1, the average time of trips done on Transmilenio is 40 minutes and the time of trips done by walk or bicycles is 30 minutes.

variable Y_i designates the average level of income for inhabitants living in zone i;

$Area_i$, represents the size in hectares of zone i;

$Dist_i$, denotes the average distance that an inhabitant of zone i takes to reach their job in t minutes;

Car_i, indicates the number of car owners in zone i;

X_i, reflects a vector of control variables to complete and fill in the missing information in the model.

What is suggested with this model is that the size of the labour market for people living in zone i depends largely on the degree of improvement to public transport in the zone, the mean income of inhabitants of each zone, the population and the area of each zone. Other variables could also explain the $ESLM$, such as the mean distance from inhabitants' houses to their jobs, the number of car owners in each zone and other control variables represented by vector X_i.

I decided not to take into consideration the variable emplo which denotes the number of jobs in each zone because these results ($ESLM$) were made directly with respect to the number of jobs on each zone; therefore a big problem of collinearity was avoided.

As previously clarified, I consider five different intervals of time of $ESLM$ so, in the first stage of our analysis (OLS estimations), three different results of this model will be presented. For the second stage (2SLS) there are three results for each interval of time in addition to the three different results for each interval of time which will depend on the size of companies on each zone (nine results are expected for this model).

Model 2 :

$$\begin{aligned} Y \quad &= \beta \quad {}_0 Cons + \beta_1 impro_i + \beta_2 ESLM_{ti} + \beta_3 Pop_i + \beta_4 Area_i + \\ &\qquad \beta_7 X_i + e \end{aligned} \tag{3}$$

Model 2 shows us the dependency of the income factor with respect to $impro_i$, $ESLM_{ti}$, Pop_i, $Area_i$ and X_i. With respect to the available data, these are the variables that best explain the income. Results of this assumption will be found on section 4 (OLS results).

The hypothesis we follow with this model and the others is that improvements to public transport in each zone have a bigger impact when considered as an IV depending on other exogenous variables (see equation 1) than when it just plays the role of an ordinary explanatory variable.

Because Income is explained by $ESLM$, I expect also to have three results for the first stage of the analysis and nine different results of the 2SLS analysis; each with respect to the commuting time and the size of enterprises.

Model 3 :

$$\begin{aligned} Emplo \quad &= \beta \quad {}_0Cons + \beta_1 impro_i + \beta_2 Y_i + \beta_3 Pop_i + \beta_4 Area_i + \\ &\qquad \beta_7 X_i + e \end{aligned} \tag{4}$$

In model 3, we consider that the number of employment on each zone of the city depends on the average income of each zone, the population and the area of that zone and, mainly, the number of Transmilenio stations in each zone; other control variables were taken in consideration as in previous models.

Furthermore, like in previous models, the lack of Transmilenio stations in a zone supposes that it does not benefit from public transport enhancements.

Contrary to previous models, employment in zone i is not explained by $ESLM$. It means that one result is expected in the first stage of our analysis (OLS). Three results are expected on 2SLS models which depend on the size of enterprises.

3 DESCRIPTIVE STATISTICS AND AVAILABLE DATA

Bogotá is the biggest city in Colombia and also one of the most densely populated cities in Latin America. It counts almost eight million inhabitants and its density is around 230 people/ha[13]. It is organized following the administrative Parisian model, which means it has 20 sub-city urban areas[14] called "Localidades", each one with its own mayor and an independent budget assigned by the main city hall with respect to the population. Because of their big physical extension, "Localidades" are subdivided in planning zones. In total, there are 112 planning zones in the city. We will develop our research question with respect to each planning zone of the city.

The data we use in our analysis come from various sources.

[13] Adapting from Suarez,2005.

[14] Urban area is composed by 19 sub city urban areas and one rural. The urban areas count 35.000 hectares. (three times Paris)

On 2005, the mobility department of the city decided to make a rather complete study about the mobility behaviour in Bogotá[15] with information about residents' travel. To estimate the job accessibility index, we use the Transport Matrix of Bogotá[16]. This matrix encloses information about all possible itineraries "from" and "to" each of the 112 planning zones of the city. After some filters of the poll and some estimations, we obtained a summary of the time that people take to get from one zone of the city to any other and the mean distance[17]. In parallel, we established the time of travel when people use public transport or their own cars. This information was compiled by the department of mobility of the city.

Table 1: Descriptive statistics for the entire sample and for each variable

Variable	Label	Obs	Mean	Std.	Min	Max
Accessible jobs on 40 minutes	ESLM_40	112	414541,30	226526,60	32151	875709
Accessible jobs on 50 minutes	ESLM_50	112	645827,90	281513,60	84793	1140640
Accessible jobs on 60 minutes	ESLM_60	112	889232,30	287322,00	184320	1320369
Income	Income	112	689235,80	566579,90	196821	3032604
Number of jobs	Employment	112	13997,52	16632,94	197	142052
Improvement degree	Improvment	112	1,92	2,41	0	10
Population on 2007	Population	112	65526,80	50644,21	776	284499
Area of the Zone (hectares)	Area	112	368,13	164,97	86	925
Mean distance of trips	Mean_dist	112	8,69	2,57	5	20
Have at less one car	Car owners	112	6778,81	6065,13	0	27569
Average cost of travel	Aver Cost	112	1878,35	704,86	332	6080
Parks on the zone	Parks	112	48,32	35,19	1	171
Establishments promoting Welfare	Social Welfare	112	63,25	63,83	0	288
Establishments provinding Health	Health	112	3,53	3,24	0	15
Schools	Schools	112	32,76	25,08	0	135
Cultural Establishments	Cultural	112	8,12	8,18	0	40
Establishments promoting Sport and Recreation	Recreation	112	11,42	9,38	0	47
Neighborhoods of each zone	Neighborhoods	112	181,16	105,28	1	526
Working Force	W_Force	112	28543,31	39705,63	0	257862
Micro-enterprises	MiE	112	2584,13	2487,72	12	13826
Small-Medium Enterprises	SME	112	123,19	195,25	0	1674
Large Enterprises	LE	112	26,79	48,35	0	427

Source: Encuesta de Calidad de Vida para Bogotá (ECV) 2007. Secretaria de Planeación del Distrito de Bogotá; Secretaria de Movilidad del Distrito; author calculations.

An emphasis has to be made regarding job accessibility. The index for job accessibility was calculated following the methodology of Prud'Homme and Lee (1999). As stated elsewhere in this research, I define "job accessibility" as what

[15] Secretaria de Movilidad del Distrito; Plan Maestro de Movilidad 2005.

[16] Obtained at the "Secretaria de Movilidad del Distrito" (Department of mobility of the city) and University of Los Andes (Bogota).

[17] Olarte, C. (2012) "Heterogeneity of social classes and job accessibility: implications of transport policies in Bogota" Working paper, Centre d' Economie de la Sorbonne (CES).

Prud'Homme and Lee call "The effective size of labour market". This theory is based on the assertion that the labour market is a function of the commuting time from inhabitants' homes to zones where people work. With that methodology we are able to estimate the number of jobs they can reach in a specific time. My calculations take into account trips made by inhabitants on private and public transport.

While the subject of this study is analysis of the endogeneity and possible causality of enhancement of the public transport system on the other variables we proposed, I suppose that private cars also benefit from those improvements. In fact, improvement to the public transport system, and more precisely the construction of Transmilenio, is also produced by an upgrading of corridors alongside Transmilenio corridors. Since the construction of TM, congestion on roads along TM corridors has decreased and the average speed has increased. Improvement of public transport produced by the construction of TM also improves travel by private car and taxi. Nevertheless, I focus my research only on travel made on public transport.

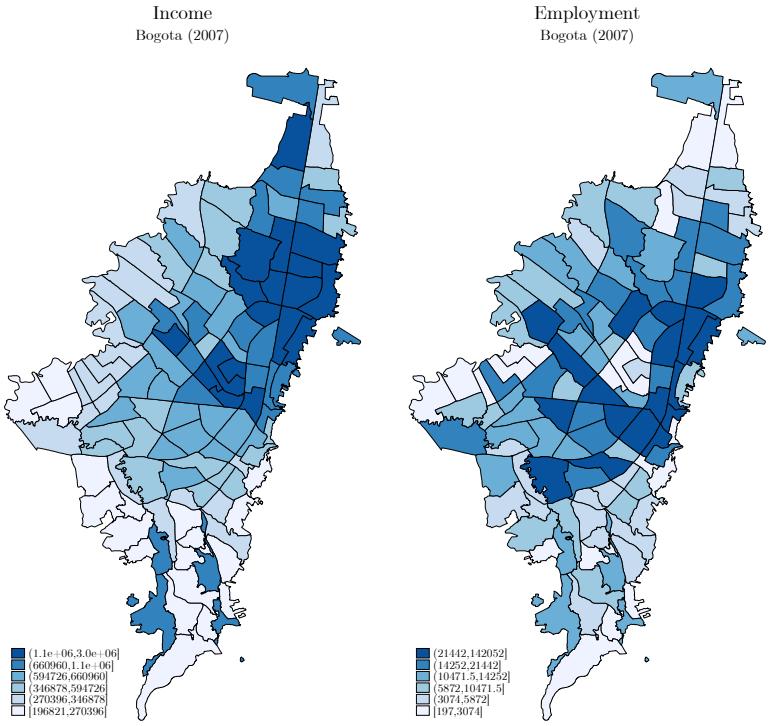
Information on socio-economic indicators such as employment, workforce, income, population and area were taken from the census conducted by the city's department of planning in 2007[18]. We could not use more recent information because the city hall has not completed censuses for subsequent years.

Data about the number of Transmilenio stations and lines within or bordering each zone were calculated by the author. To make those calculations, maps of the city were created, featuring the boundaries of each planning zone and the streets passed by Transmilenio. Stations within each zone were considered stations of only this zone. Stations on the border between two or more zones were considered stations belonging to both zones because both zones benefit from the station.

To complete our model I took into consideration other control variables such as the total number of men, the number of neighbourhoods in each zone and the number of cars per household. Additionally, variables describing some infrastructural characteristics and amenities of zones were taken into consideration: the number of parks, the number of schools and the number of establishments providing welfare, sport, recreation and cultural events in each zone, as well as the number of establishments providing health care, also make up part of the model.

[18] Encuesta de Calidad de Vida para Bogotá (ECV) 2007. Secretaria de Planeación del Distrito de Bogotá.

Figure 1: Income and Employment configuration of Bogota



Author calculations based on data from Planning Department
(Secretaria de planeacion del Distrito)

Figure 1 shows a disproportionate configuration of employment and income in the city. It is easy to identify the axes where jobs and income are concentrated[19]. Two different "job centres" are connected on the same two axes. The city's highest incomes are concentrated in two zones. These can be distinguished and they are also reliable on the same axis than the two jobs centres. This axis has the particularity that it benefits from the presence of TM. A third job axis is also placed alongside another Transmilenio axis. Regarding the other zones with high levels of income, they are not very far from the TM corridor that goes from the centre to the west of the city.

Additionally, jobs are condensed far from the boundaries of the city (with the exception of the East border). Even if we see that job centres are placed at the centre-east border of the city, this border was always considered as the "centre". Executive and almost the 100% of administrative buildings are place on that zone. Furthermore, is in that place where Bogota was founded and it could not be expanded to the East because of the presence of a big chain of mountains

[19] Configuration of $ESLM$ depending on travel time can be found on annex 4

(Los Cerros Orientales). The expansion of the city took place essentially to the south and the north of the city and more recently to the west of the city.

Following the results of a poll made by the administration in 2005, most work travel moves from the "far" north, the south and the west to these four job centres. These are the zones with the lowest income in the city, with the exception of one zone in the north, where people with high purchasing power settled a few years ago, seeking to be far from the noise and pollution of the city.

The "centre" of the city condensed more than 50% of jobs in the city and according to the poll, most of the people commuting to this zone of the city come from the north of the city. Others travel from the north-west, the west and the south and south-west of the city.

Apparently, this configuration of Bogotanians' jobs and income of can be directly related to the enhancement of the public transport system.

The next section will focus on analysis of this possible strong relationship. Results of the models proposed in section 2 will provide some elements supporting knowledge of the existence and magnitude of this relation.

4 RESULTS AND DISCUTION

As stated above, analysis of the correlation matrix and the variance inflation factor (VIF) of all models is recommended.

The correlation matrix[20] shows important relationships between some variables[21], in particular the variables "Population" and "Schools". In effect, the population has correlations above 0.50, with almost all the variables representing the characteristics or the amenities of each zone. These correlations can be interpreted as reasonable. In effect, the higher the population is, the more zones will need schools, parks, recreational facilities and sporting establishments. Likewise, if the population is bigger, it can be expected that the number of neighbourhoods will be greater as equal as the area of the zone. In despite of this, correlations just represent some control variables and not make part of the core of our analysis, it let infer that these correlations will not affect the estimations.

Nevertheless, the greater correlation is between the number of schools in zones and the number of establishments promoting sport and recreation (0.81). This is predictable if we consider that the most of the schools in Bogotá are public. They usually use the infrastructure of the zones (parks, and recreation establishments) for students' recreation and sporting activities.

[20] Annex 1.
[21] We consider as important, correlations upper 0.50.

Another important correlation that relates to the goal of this study is between employment and improvement of public transport (number of stations); they have a correlation index equal to 0.56. It can be also awaited but it seems not to be a problem of multicollinearity if we consider that VIF is smaller than 10. It could also represent an additional argument to suggest that transport improvements directly affect the number of jobs (model 3).

Other correlation indices above 0.50 are found. The big indices are principally between amenities variables, which is not very surprising and will not be a source of evils. None of those variables explain each other.

On the other hand, and as indicated previously for some relations between variables, regarding (VIF), no problems have been detected in the models[22]. On every model it is found that VIF < 10, which means that multicollinearity is low and dependent variables are uncorrelated with predictor variables.

The following section will present the results of OLS estimations and 2SLS estimations.

4.1 OLS Results

As noted previously, for the first stage of the analysis, each model – depending on the effective size of the labour market – is regressed three times, in accordance with the three intervals of times. Table 2 shows results for the three models. There are three different results for models 1 and 2 and one result for model three.

4.1.1 Model 1: Effective Size of Labor Market

Regarding model 1, results show a not negligible R^2 for each travel time, which goes from 0.583 when travel time is 40 minutes to 0.615 when travel time is 60 minutes. Taking into account the analysis of the correlation matrix and also the analysis of the VIF, it can be said that the model explains in an acceptable proportion the dependent variable.

Regarding regressors, table 2 suggests that *improvement* has a positive and big effect on the level of the effective size of labour market. In effect, for a time period of 40 minutes, the number of Transmilenio stations seems to boost the size of the labour market of inhabitants of zone i on 15,200 jobs. Furthermore, for time intervals of 50 and 60 minutes, the presence of Transmilenio improvements in the zone has a positive relationship with the size of the labour market in the

[22] Annex 2.

order of 20,314 and 21,388 jobs respectively. These three results are significant at the 0.05 level, which is not negligible.

Regarding income on each zone, it is also shown that it has a positive impact on the size of the labour market, which was predictable. The higher the income is, the higher will be the size of the labour market, because inhabitants may choose to live in a neighbourhood that is closer to their workplace. Nonetheless, in model 1, income seems not to have a significant impact on $ESLM$.

Another variable that appears to significantly impact the $ESLM$ is the mean distance between inhabitants' homes and their jobs. This variable has a negative impact on the number of jobs reachable within these intervals of time. For the three different time periods, estimators of mean distance between reachable jobs and houses are significant at the 0.01 level.

Additionally, its influence on the level of $ESLM$ looks to be very important. Actually, an additional kilometre of mean distance reduces the $ESLM$ by 5% and 10% for any interval of time. It may seem strange but it can be considered normal. In effect, concerning the $ESLM$, results confirm the fact that, when individuals live far from their jobs, the number of reachable jobs relevant to their skills in the zone where they live may decrease. This means that the size of the labour market for people decreases if their jobs are far from their houses.

The variable $Area$ is another variable that seems to be significant from a statistical point of view. In effect, table 2 shows that for any interval of time, it is significant at a 0.05 level. Nevertheless it has always, a negative sign which can be estranged. Actually, regarding the area of the zone and its negative relationship with the size of the labour market for each population living in each zone, we can suggest that it could be because jobs and income are concentrated, in a big proportion, on small zones which are placed close to the job centres, while the biggest zones are placed at the periphery of the city.

People living in the biggest zones are those with smaller levels of income and are farther from job centres. In parallel, the bigger the area of each zone, the higher the number of neighbourhoods will be; that may also be the reason for the negative sign of the neighbourhoods parameter.

Table 2: OLS results for each interval of time

	SLM			Income			Employment
	slm 40	slm 50	slm 60	income 40	income 50	income 60	
Intercept	711,644.313***	1045928.176***	1313650.550***	-136,230.614	-134,996.960	-209,878.949	-3,632.736
	(79,704.801)	(97,313.955)	(97,115.364)	(312,658.436)	(344,254.457)	(394,996.977)	(5,667.080)
Improvment	15,220.654**	20,314.682**	21,388.184**	25,161.388	26,034.941	25,842.513	3,274.915***
	(7,523.774)	(9,185.999)	(9,167.252)	(22,197.682)	(22,362.994)	(22,434.205)	(534.947)
Job_Accessibility				0.450	0.307	0.301	
				(0.292)	(0.241)	(0.242)	
Income	0.053	0.054	0.052				0.010***
	(0.034)	(0.042)	(0.042)				(0.002)
Population	-0.362	-0.324	-0.441	-3.854*	-3.949*	-3.918*	-0.037
	(0.723)	(0.883)	(0.881)	(2.069)	(2.074)	(2.077)	(0.051)
Area	-262.545**	-385.040**	-388.453**	336.586	338.087	337.181	21.047**
	(121.573)	(148.432)	(148.130)	(360.452)	(365.529)	(365.922)	(8.644)
Mean_dist	-35,218.349***	-45,881.804***	-50,838.129***	6,305.226	4,437.753	5,676.773	-408.308
	(7,161.193)	(8,743.313)	(8,725.470)	(23,276.488)	(23,691.774)	(24,299.335)	(509.167)
Car owners	-3.859	-1.873	-0.770	8.649	7.539	7.201	0.229
	(3.465)	(4.231)	(4.222)	(10.107)	(10.101)	(10.099)	(0.246)
Aver Cost	59.329**	75.797**	86.296***	230.366***	235.832***	233.249***	0.221
	(26.480)	(32.331)	(32.265)	(75.460)	(75.799)	(76.579)	(1.883)
Parks	-322.012	-781.846	-151.965	4,351.384**	4,478.577**	4,287.399**	-89.009*
	(713.883)	(871.601)	(869.823)	(2,031.430)	(2,043.390)	(2,039.829)	(50.758)
Social Welfare	-933.344**	-1,140.505**	-1,007.571**	-2,465.939**	-2,558.896**	-2,607.019**	-7.563
	(385.331)	(470.462)	(469.502)	(1,126.882)	(1,129.365)	(1,121.241)	(27.397)
Health	-2,964.688	-3,710.037	-2,941.590	-18,035.172	-18,382.453	-18,646.856	234.778
	(5,759.372)	(7,031.788)	(7,017.438)	(16,677.106)	(16,739.421)	(16,733.046)	(409.496)
Schools	1,053.693	1,486.035	1,243.575	-4,910.330	-4,925.664	-4,847.548	-32.605
	(1,644.851)	(2,008.247)	(2,004.149)	(4,769.123)	(4,790.947)	(4,789.328)	(116.950)
Cultural	-1,027.805	85.714	-1,576.168	17,458.404*	17,100.412*	17,612.816*	984.440***
	(3,163.854)	(3,862.843)	(3,854.960)	(9,036.689)	(9,074.786)	(9,075.872)	(224.953)
Recreation	5,871.716*	5,689.958	5,259.589	15,048.130	16,084.718*	16,256.850*	16.487
	(3,270.352)	(3,992.870)	(3,984.722)	(9,549.043)	(9,510.627)	(9,496.559)	(232.525)
Neighborhoods	-649.501	-765.556	-1,110.307*	638.618	583.495	683.448	-51.926
	(516.125)	(630.153)	(628.867)	(1,512.266)	(1,517.503)	(1,530.195)	(36.697)
Obs	112	112,00	112	112	112	112	112
R^2	0.583	0.598	0.615	0.436	0.432	0.432	0.609

Author calculations

The estimator for the variable Average cost, which represents the average cost of travel that people have to pay to reach their job, appears with a positive sign on table 2. Moreover, this estimator is significant at a 0.05 level with any travel time. It suggests that the more I pay for transport to reach my job, the higher will be the size of my labour market. It could also be expected if we consider that the price to be paid for use of the public transport we consider in this study (Transmilenio) is higher with respect to the other kind of public transport services, excepting taxis. Additionally, all Transmilenio lines were constructed on an axis passing by job centres, which makes Transmilenio the more expensive but, at the same time, the quicker transport system to reach job centres.

Regarding amenities in the zones, table 2 shows that four variables seem to have a negative effect on the $ESLM$ and two variables have positive impacts. Among these six amenities, just one has a significant influence on the dependent variable. $SocialWelfare$ has, for the three intervals of time, a negative influence on dependent variable with a significance level of 0.05. It may be because this variable denotes the number of establishments like nursing homes, rehabilitation centres, orphanages and establishments promoting the welfare of inhabitants

with some problems interacting with society. In that vein, this kind of establishment may not have a positive effect on the labour market because it requires a lot of space in the zone and employs fewer people than companies carrying out other activities. As with $SocialWelfare$, estimators of variables $Parks$ and establishments promoting cultural activities, $Cultural$, have a negative incidence on the $ESLM$ but do not have significant influence on the dependent variable. Reasons may be the same as for $SocialWelfare$. $Parks$ and $Cultural$ take a lot of vital space in the zone and avoid the construction of roads, lines of transport systems and job centres.

Regarding establishments providing health services, $Health$, results show that they has a negative impact on $ESLM$. This seems to be counterintuitive because hospitals and health centres are supposed to create jobs. It suggests that the decision to construct hospitals in the city depends on the available area in each zone. In effect, zones where job centres are situated do not have as many areas available for construction of hospitals as the peripheral zones do. In addition, the expensiveness of domiciliary public services on jobs centres could also influence this relationship.

In summary, if a comparison of estimators of all variables is made, it can be said that the variable that most influences $ESLM$ is Improvement. In addition, its influence increases with respect to the commuting time, meaning the greater individuals' commuting time, the greater will be the influence of public transports improvement on the effective size of the labour market.

In opposition, the other variable that has a significant but negative impact on the $ESLM$, with a statistical significance of 0.01, is the mean distance that people have to commute to their jobs. As with $improvement$, $Mean\_dist$ increases with commuting time, which is expected. The higher the mean distance commuted by people, the lower the $ESLM$ of inhabitants living on origin zones.

4.1.2 Model 2: Level of Income

Results for model 2 are also divided into three because one of the regressors is $ESLM$ and it varies with respect to commuting time. On the other hand, while influences of some estimators are not different from model 1, there are also five big differences that have to be remarked upon.

First of all, the variable Improvement has a big influence on the level of income of inhabitants of each zone. This suggests that the higher the presence of public transport improvements, the greater will be inhabitants' income. This relationship can be interpreted from two different points of view. In effect, it is not false to suggest that the presence of Transmilenio can boost the income of inhabitants. On the other hand, it can also be said that people with greater levels of income choose to live in zones where there is more improvement of

urban transport. Nevertheless, it is clear that the impact of improvement of urban transport on the level of inhabitants' income cannot be denied. However, while the improvement estimator is the one that has the largest influence on the level of income, it also has a problem in the OLS results: it is not statistically significant. It represents another additional reason to carry out the 2SLS analysis in section 4.2.

The other difference between model 1 and model 2 is that the estimators for $Area$, $Mean\_Dist$, $Parks$, $Cultural$ and $Neighbourhoods$ are not negative but positive. First of all, regarding the area of the zone, it indicates that the bigger the zone, the greater the average income. This relationship is expected. In effect, as suggested by Anas (1990), Glaeser, Kahn and Rappaport (2000) and other researchers, rich people sometimes prefer to live in suburbs or on zones far from the city to avoid the noise and the congestion of job centres. Additionally, in suburbs or zones far from the centre, rich people find more space to live or to construct bigger houses than those that they find downtown or near job centres. In parallel, amenities they prioritize, aside from housing space, are safety, calm (no noise or pollution), large green spaces to do sport and for their children to play in and proximity to nature, in spite of the need to consider proximity to their jobs or to the centre of the city. This is the reason for the large and positive influences of estimators for $Parks$, $Recreation$, $Cultural$ and $Neighbourhoods$, of which $Parks$ and $Cultural$ are those that are statistically significant, at 0.05 and 0.1 respectively.

Regarding the $Cultural$ estimator and establishments promoting cultural activities, it could be interpreted that cultural manifestations like theatre, opera and concerts, among others, are not affordable for people with a limited income. Those kind of cultural expressions are revealed to be expensive, preventing accessibility by people with a limited budget; that could be the reason for the positive and statistical significance of this estimator on income.

In contrast, $SocialWelfare$, $Health$ and $Schools$ have a negative and not negligible influence on the level of income, but only $SocialWelfare$ is statistically significant (0.05). It can be interpreted as meaning that if there are more hospitals, schools, rehabilitation centres or nursing homes, this may lead to congestion; the calm of the zone can be affected and people with bigger incomes may not be incentivized to live near those kinds of establishments. The easiest and best solution for people with high incomes is to live in distant and expensive suburbs with small density and with fewer establishments promoting health, social welfare or schools. In opposition, people with low incomes also decide to live far from the city and job centres, but in suburbs or zones with high density and important concentrations of hospitals, schools and establishments providing social welfare. Likewise those variables, the negative and statistically significant (0.01) impact of the variable $Population$ is explained.

Finally, the other variable that reveals a statistically significant impact on the level of income is $Aver\_Cost$; this is also expected because if rich people

decide to live far from their jobs, they will be pushed (voluntarily) to expend more money to commute from their houses to their jobs.

As in model 1, model 2 reveals that the variable with more influence on the dependent variable is the level of improvement of public transport in each zone. Nevertheless, with respect to the level of income, this variable is not significant from a statistical point of view, and this leads us to the 2SLS analysis in section 2.2.

4.1.3 Model 3: Number of jobs

Model 3 shows the influence of some explanatory variables on the number of jobs in each zone of the city. This model is measured with respect to the same independent variables as in model 1, but there is just one result because it does not depend on $ESLM$. Subsequently, even if the results are similar, there are also some important differences that should be clarified.

The first difference is that income is now statistically significant, at 0.01, which was not the case in model 1. This may suggest that the number of jobs will be bigger in zones where income is higher, but it also may be proposed that enterprises or companies settle in zones or near zones where income is higher (Zenou, 2000, 2008; Ross, 1998; Kain, 1968).

Secondly, table 2 shows that the area of the zone has a positive and a statistical significance (0.05) to the number of jobs in each zone. The bigger the zone, the more jobs there will be in the zone. This result is the opposite from what was found in model 1, because model 1 tries to determine the impact of the area of the zone on the inhabitants' number of reachable jobs. What was suggested in model 1 was that area has a negative impact on $ESLM$; this is very different if the dependent variable is the number of jobs in each zone. $ESLM$ takes into account reachable jobs for each inhabitant with respect to their skills and employment refers to all kinds of jobs in each zone. So, this could be the reason for this differing relationship between models.

Another regressor that is significant from a statistical point of view is the number of parks in each zone. This variable has a negative influence on the number of jobs in each zone, which is logical because the more parks in each zone, the less available space there will be for the settlement of enterprises in the zone will be.

Fourthly, establishments promoting cultural activities seem to boost – and to create a significant amount of – the number of jobs in each zone. This result suggests that cultural activities are an important source of employment for the city, equal to establishments providing health care. *Schools* seems to have the opposite effect to $Health$ or $Cultural$.

Aside all other results which are similar and could be read like those in model 1, the most important to note is that the variable with the bigger impact on the number of jobs on each zone is, once again, the number of Transmilenio stations. In addition to this, the estimator of this variable is significant at 0.01 level, which demonstrates the great dependence of employment on improvements to public transport. It could signify that the number of employment grows by 3.274 if an additional Transmilenio station is built in a specific zone. However, it could also denote that enterprises decide to settle in zones with high levels of improvement to public transports and great levels of accessibility (Fernandez, 2008; Kawabata, 2003).

On the three precedent models, table 2 shows that the variable *improvement* has always a big and almost the greater influence on the dependent variable, regardless of commuting time, when the variable ESLM is part of the model. Nonetheless, regarding the statistical significance, this variable is not significant in model 2.

2SLS analysis is then necessary to verify if the results of these models are not forgetting instruments that can determine or influence the number of Transmilenio stations on each zone. In other words, 2SLS analysis will let us verify if there exists an endogenous relationship between dependent variables of the three models and the variable *improvement* and, therefore, a causal relation between the number of stations in each zone and the dependent variables of each model.

4.2 2SLS Results

As defined in section 2.2, the third stage of this analysis is focused on the identification of endogeneity between the level of public transport and dependent variables. As for OLS results, it is essential to clarify that models 1 and 2 have three results depending on the commuting time.

Additionally, the three models will also be regressed three times according to the definition of the Instrument Variable (IV). In effect, it is suggested that the decision to construct a Transmilenio station in a zone depends directly on the density, the workforce and the number of enterprises on each zone[23]. In that vein, the size or the number of employees of enterprises based in each zone should be taken into account and should be differentiated. For that reason, this study proposed three different regressions of each model regarding the definition of the IV and in accordance with the size of companies established in each zone.

[23] See section 2.2 for definition of what it was called Microenterprises, Middle size enterprises and Large enterprises

4.2.1 Enhancements of PT defined by the number of microenterprises on zone i.

Results presented in this subsection correspond to models 1, 2 and 3 when the VI is defined by the density, the workforce in each zone and the number of microenterprises in each zone.

First of all, table 3 shows that every estimator has the same relationship (sign) and the same statistical significance for model 1 than on OLS results. Improvement is the variable that has the higher positive and statistical influence on the size of the labour market in each zone. In parallel, the mean distance between houses and jobs is the variable that has the higher negative and statistical impact on the dependent variable.

Table 3: 2SLS results for each interval of time when:
$impro = \pi_0 + \pi_1 WF_i + \pi_2 MiE_{si}$

	SLM			Income			Employment
	40 minutes	50 minutes	60 minutes	40 minutes	50 minutes	60 minutes	
Intercept	703,963.433***	1031718.345***	1302852.910***	-121,333.320	-108,458.695	-159,857.237	-10,059.739
	(76,209.254)	(93,583.480)	(93,355.711)	(314,339.274)	(323,163.063)	(372,672.907)	(6,948.676)
Improvment	22,117.543*	32,521.548**	31,071.155**	40,792.818	41,285.053	43,053.538	7,588.386***
	(12,173.235)	(14,948.495)	(14,912.112)	(38,949.393)	(36,786.388)	(36,908.604)	(1,109.942)
SLM				0.394	0.258	0.243	
				(0.306)	(0.236)	(0.238)	
Income	0.047	0.043	0.044				0.007**
	(0.033)	(0.041)	(0.040)				(0.003)
Population	-0.441	-0.423	-0.578	-3.634*	-3.729*	-3.694*	-0.012
	(0.671)	(0.824)	(0.822)	(2.060)	(1.922)	(1.925)	(0.061)
Area	-0.020**	-0.029**	-0.028**	0.025	0.024	0.024	0.002**
	(0.010)	(0.012)	(0.012)	(0.031)	(0.029)	(0.029)	(0.001)
Mean_dist	-36,208.470***	-47,168.293***	-52,548.061***	6,706.096	4,535.453	5,150.545	-177.267
	(6,614.546)	(8,122.534)	(8,102.765)	(23,290.065)	(22,132.616)	(22,823.788)	(603.107)
Car owners	-3.272	-0.892	0.086	9.093	8.080	7.885	0.484
	(3.307)	(4.061)	(4.052)	(10.239)	(9.555)	(9.574)	(0.302)
Aver Cost	59.490**	76.436**	86.289***	236.064***	241.762***	240.669***	0.977
	(24.888)	(30.562)	(30.487)	(75.742)	(70.807)	(71.623)	(2.269)
Parks	-324.728	-774.181	-159.629	4,440.929**	4,544.679**	4,388.467**	-62.548
	(671.884)	(825.060)	(823.052)	(2,038.148)	(1,907.578)	(1,905.703)	(61.262)
Social Welfare	-905.372**	-1,094.140**	-966.178**	-2,472.263**	-2,569.331**	-2,614.283**	4.153
	(362.732)	(445.428)	(444.344)	(1,130.620)	(1,053.969)	(1,047.196)	(33.074)
Health	-3,834.870	-5,175.738	-4,215.141	-19,269.376	-19,584.524	-19,986.315	-174.459
	(5,490.552)	(6,742.291)	(6,725.881)	(17,005.259)	(15,888.868)	(15,885.761)	(500.622)
Schools	763.630	979.882	839.208	-5,357.926	-5,340.447	-5,325.691	-191.838
	(1,588.269)	(1,950.364)	(1,945.617)	(4,889.940)	(4,567.202)	(4,571.254)	(144.817)
Cultural	-353.538	1,227.071	-595.260	18,263.733*	17,930.487**	18,462.362**	1,309.516***
	(3,065.944)	(3,764.920)	(3,755.757)	(9,293.874)	(8,707.936)	(8,684.017)	(279.549)
Recreation	5,340.337*	4,722.659	4,513.505	13,805.633	14,835.026	14,872.610	-385.920
	(3,202.302)	(3,932.365)	(3,922.794)	(9,940.897)	(9,245.787)	(9,239.208)	(291.982)
Neighborhoods	-699.074	-834.504	-1,196.246**	660.919	603.795	677.960	-52.580
	(482.502)	(592.503)	(591.061)	(1,516.013)	(1,415.777)	(1,432.348)	(43.994)
Endogeneity - Wu - Hausman F Test	0.44933	0.94049	0.58849	0.22424	0.20373	0.25837	48.18274***
Sargan	3.337	4.646*	1.956	4.002	4.158	4.459	2.690
Heterocedasticity	19.211	18.153	15.928	36.676 ***	36.388 ***	34.986 ***	19.623
R^2	0.576	0.586	0.605	0.432	0.428	0.427	0.346

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Author calculations

Although there is a similarity between OLS and 2SLS results for model 1, the impact of the number of stations on the size of the labour market is 50% higher with respect to OLS model results. .Following this observation, it is suggested that when improvement is treated as an IV depending on the number of microenterprises, its impact on the dependent variable increases by 50%.

Regarding the endogeneity analysis, table 3 reveals that even if the number of stations in each zone remains significant and greater than in OLS regressions, and even if there is no problem regarding the choice of instruments that explain IV, the variable improvement is not endogenous to the size of the labour market. This result is counterintuitive to the objective that this investigation tries to demonstrate but one fact is salvageable, which is that when improvement is considered as a IV, its statistical significance and its impact on $ESLM$ increase by 50%.

Concerning model 2, 2SLS results are not significantly different from those found in OLS estimations. Similarly, relationships between variables are respected and every estimator preserves the same sign, the same statistical significance and almost the same level. It means that the number of stations in each zone is not yet statistically significant but their effect on $ESLM$ is almost 70% higher than in the OLS models.

As in the previous model, selected instruments do not have problems of over-identification but the analysis of endogeneity displays that "improvement" is not endogenous to the income of each zone's inhabitants. It was not what the study had expected and it will be explained at the end of all 2SLS results' discussion.

Results for model 3 are slightly different. Variables such as *income*, *Population*, $Aver\_Cost$ and Car_owners preserve the same impact and the same statistical inference on the dependent variable as in OLS results. Likewise, the mean distance between homes and jobs has the same relationship as in the OLS model but its influence on independent variable increases by 100%. In addition, even if the variable *Area* has the same statistical influence (0.05) on the number of jobs as in the OLS model, its "real" impact is marginally (0.002) contrary to the number of parks on each zone that is no longer statistical significant but its real influence on employment is also negligible. The other group of variables does not have a significant influence on the number of jobs in each zone, with the exception of two variables that are very significant and that have a notable influence on the number of jobs.

The two regressors that have a significant effect on employment in each zone are the number of establishments promoting cultural activities and the number of stations in each zone. Regarding *Cultural*, its statistical significance is high (0.01). This result could signify that the implantation of one additional establishment promoting cultural activities could raise by 1.309 the number of jobs in each zone. On the other hand, this result could be interpreted differently; the target population of that kind of establishment could be employees, but this study does not have more enough information to make this statement.

With respect to the number of stations, as in the OLS results, it can be observed that this is the regressor with the greater effect on the explained variable. When *improvement* is considered as endogenous and when it depends on the number of microenterprises, its influence on the number of jobs in each zone is double that of the OLS model. In addition, its statistical significance is still very high (0.01). In effect, an additional station in each zone is supposed to boost the number of jobs by 7,560. Besides, it can also be supposed that Transmilenio stations were built with the aim to be close to employment centres. Neither hypothesis moves away from the goal of this analysis, which is to demonstrate the causal relation between the number of jobs in each zone and the level of improvement to public transports.

Regarding this statement, endogeneity test reveals that *improvement* is endogenous to the number of jobs on each zone. In other words, the number of Transmilenio stations is endogenous with respect to the number of jobs. This result implies that the hypothesis is verified, which enables us to prove the following hypothesis:

$$H_0 : \rho(improvement) = 0(exogeneity)$$
$$H_1 : \rho(improvement) = 0(endogeneity)$$

The p-value of Wu-Hausman F test is statistically significant (0.01), which leads to rejecting the null hypothesis of exogeneity, H_0. The endogeneity in the relation between *improvement* and the number of jobs in each zone is demonstrated.

This result is even truer if, the test of over-identifying restrictions, "Sargan N*R-sq test"[24], and heterocesdasticity test (Pagan-Hall [25]) are taken into consideration. In fact, those tests demonstrate that instruments have no problems of over-identification and heteroscedasticity (no rejection of null hypothesis). This means that instruments are exogenous and the residuals of the main model are uncorrelated with the set of instrumental variables. Instruments were well chosen.

Considering results for models 1 and 2, OLS results appear to be consistent. In effect, *improvement* is not endogenous when it is regressed depending on density, working force and microenterprises. By contrast, OLS results are not consistent for model 3. In fact, suspicions of endogeneity of *improvement* are corroborated, which represents an undeniable causality relation from enhancement of urban transport to the number of jobs when the number of stations are defined by microenterprises.

[24] The joint null hypothesis is that the instruments are valid instruments, i.e., uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation. Under the null, the test statistic is distributed as chi-squared in the number of (L-K) overidentifying restrictions. A rejection casts doubt on the validity of the instruments.

[25] H_0 : Disturbance is homoscedastic

4.2.2 Improvement regressed with respect to "small and middle size" enterprises

Results presented in this subsection correspond to models 1, 2 and 3 where the variable suspected to be endogenous is defined by the density, the work force in each zone and the number of middle-size enterprises in each zone.

Results for model 1 express exactly the same relationships as in the previous subsection. This means that there is no evidence to reject the use of instrumental variables on the model. In effect, Sargan test shows that there is no over-identification of instrumental variables likewise with the heteroscedasticity test that confirm that the variables are homoscedastic. Nevertheless, as in the previous model, Endogeneity – Wu-Hausman F test allows to accept null hypothesis of exogeneity; the number of stations on each zone of the city are not endogenous on the effective size of labour market. Taking into account those previous observations, it can be confirmed that OLS results are consistent for model 1.

In contrast, the results for model 2 suppose different interpretations. The variable *Population* continues to have the same impact on the dependent variable but it ceases to have statistical significance; also, $Social\_Welfare$ preserves its influence on Income but is no longer significant at 0.05; rather it is at 0.1. On the other hand, establishments providing health services increase their negative influence on the level of income by 70% more than when *improvement* is defined by SME. Additionally, this variable has a statistical significance of 0.1, which is not very representative but suggests that if the number of middle-size enterprises is taken into account at the moment to define the variable improvement, an additional establishment providing health in the zone will lower the income of inhabitants by 34.500 COP[26]. It represents a decrease of 8%[27] of the minimum wage established by the government on 2007. These results could suggest that inhabitants with higher incomes prefer to live in zones with few promoters entities of health. The income of inhabitants living in zones with several promoter entities of health is 34.500 COP lower than that of inhabitants of zones with few of those kinds of establishments, which represents 8% of minimum wage for that year.

Regarding the number of Transmilenio stations in a zone and their effect on the level of income, table 4 reveals that this influence is notable. In effect, results suggest that one additional Transmilenio station will boost the income of inhabitants by 197.000 COP[28], which represents 28% of the average wage of a citizen and 45% of the minimum wage decreed by the Colombian government

[26] COP: Colombian Pesos. The exchange rate on 2007 was: 1 USD = 2078 COP; 34.500 COP = 16.6 USD

[27] The minimum wage represents the minimum wage that a worker have to be paid monthly; on 2007, the minimum wage in Colombia was 433.700 COP = 208 USD.

[28] 197.000 COP = 95 USD on 2007.

in 2007[29]. But this result could also be interpreted as meaning that people that live in a zone with one more station than another zone will have an income 197.000 COP higher. This has a statistical significance of 0.01. In addition, results show that the variable *improvement* is endogenous and that there is no evidence against the veracity of instruments used on the model. On the same way, instruments do not present problems of heteroscedasticity.

Previous results affirm the hypothesis that when *improvement* depends on the density of the zone, the workforce of the zone and the number of "middle-size" enterprises, there is a causal relationship between the enhancement of public transport and the level of income of inhabitants.

Table 4: 2SLS results for each interval of time when:
$impro = \pi_0 + \pi_1 WF_i + \pi_2 SME_{si} + u$

	SLM			Income			Employment
	40 minutes	50 minutes	60 minutes	40 minutes	50 minutes	60 minutes	
Intercept	708,406.951***	1032649.269***	1295251.120***	-39,144.142	64,556.348	139,053.591	-15,271.339
	(76,485.215)	(94,123.698)	(94,719.511)	(367,614.009)	(407,977.334)	(472,640.016)	(9,751.496)
Improvment	19,046.838	31,878.230*	36,324.391**	193,049.558***	196,398.682***	197,621.601***	11,189.875***
				(51,658.167)	(52,688.143)	(53,398.809)	
SLM				-0.042	-0.129	-0.159	
	(13,579.695)	(16,711.348)	(16,817.133)				(1,731.346)
Income	0.049	0.044	0.040				0.005
	(0.033)	(0.041)	(0.041)	(0.364)	(0.304)	(0.307)	(0.004)
Population	-0.453	-0.425	-0.558	-2.948	-2.999	-3.039	0.002
	(0.670)	(0.824)	(0.829)	(2.410)	(2.423)	(2.430)	(0.085)
Area	-0.020**	-0.030**	-0.028**	0.028	0.025	0.025	0.002*
	(0.010)	(0.012)	(0.012)	(0.036)	(0.037)	(0.037)	(0.001)
Mean_dist	-36,304.776***	-47,188.469***	-52,383.304***	-3,876.159	-8,445.628	-10,747.558	-64.314
	(6,597.005)	(8,118.360)	(8,169.750)	(27,272.652)	(27,954.481)	(28,905.269)	(841.086)
Car owners	-3.463	-0.932	0.413	16.128	16.175	16.338	0.707*
	(3.319)	(4.084)	(4.110)	(12.020)	(12.106)	(12.157)	(0.423)
Aver Cost	59.005**	76.334**	87.119***	263.892***	272.159***	276.431***	1.546
	(24.830)	(30.556)	(30.750)	(88.634)	(89.315)	(90.516)	(3.166)
Parks	-341.290	-777.651	-131.297	4,736.085**	4,665.679*	4,746.860**	-43.124
	(670.627)	(825.282)	(830.506)	(2,382.514)	(2,402.630)	(2,403.754)	(85.502)
Social Welfare	-914.176**	-1,095.984**	-951.116**	-2,214.583*	-2,327.799*	-2,343.030*	14.479
	(362.040)	(445.532)	(448.352)	(1,322.035)	(1,328.025)	(1,321.256)	(46.158)
Health	-3,532.719	-5,112.437	-4,732.048	-33,668.621*	-34,241.089*	-34,309.575*	-528.838
	(5,506.433)	(6,776.288)	(6,819.183)	(20,008.288)	(20,151.585)	(20,177.473)	(702.044)
Schools	878.794	1,004.009	642.191	-10,110.038*	-10,031.781*	-10,044.785*	-326.908
	(1,599.799)	(1,968.733)	(1,981.195)	(5,765.620)	(5,802.116)	(5,819.265)	(203.967)
Cultural	-592.708	1,176.964	-186.098	28,044.848**	28,278.183**	28,076.871**	1,590.028***
	(3,093.157)	(3,806.479)	(3,830.575)	(10,974.733)	(11,094.413)	(11,070.517)	(394.363)
Recreation	5,621.119*	4,781.483	4,033.156	1,418.761	1,867.156	1,960.132	-715.235*
	(3,240.725)	(3,988.079)	(4,013.324)	(11,786.867)	(11,832.242)	(11,853.673)	(413.177)
Neighborhoods	-696.189	-833.899	-1,201.182**	186.329	110.003	26.772	-55.965
	(481.055)	(591.992)	(595.740)	(1,773.443)	(1,784.926)	(1,809.478)	(61.332)
Endogeneity - Wu - Hausman F Test	0.10503	0.61123	0.99825	23.88448***	23.52866 ***	23.15102***	432.31390 ***
Sargan	4.229	5.052*	1.694	0.898	0.907	0.973	0.496
Heterocedasticity	19.863	18.270	15.567	17.900	17.902	17.916	9.486
R^2	0.579	0.5869	0.598	0.104	0.093	0.089	-0.270

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Author calculations

Regarding the variable *Mean_Distance* and comparing its estimator with the one used when the instrumental variable was the number of microenterprises

[29] See table 1: Average wage per month for 2007 at the city Bogota was 687.235 COP which on dollars of 2007 is 331 USD.

on each zone, table 4 reveals a change of sign. In effect, there is a negative influence of the mean distance on the level of income and, even if it is not significant from a statistical point of view, its influence is bigger than in precedent models. It suggests that if the distance between homes and jobs increases, it has a negative impact on the income of inhabitants. In other words, what this result suggests is that SME are not placed close to zones where inhabitants have a small income or that SME are situated close to people that have higher incomes. On the other hand it could also reveal that people with greater incomes can decide to live closer to SME than people with modest incomes, and thus the negative influence of $Mean\_Distance$ on $Income$.

Concerning the variable $Schools$, it is also seen that its influence on the dependent variable is doubled with respect to the previous model and their relationship remains negative. In addition, this variable is statistically significant at 0.1 which convert it on a not negligible variable on the model.

Finally, the two variables that have the biggest positive influence on the level of income are the number of establishments promoting cultural events and the number of Transmilenio stations. Regarding $Cultural$, its influence on the level of income increases by 60%. It means that, if an additional establishment promoting cultural activities is built in a zone, this could boost the income of inhabitants by 28.000 COP. This has a statistical significance of 0.05.

Results for model 3 do not differ from the model with "microenterprises" as an instrumental variable. Actually, all the variables have the same influence on inhabitants' income. Statistical significance remains the same as in the previous model. The variables that most significantly affect the level of income are $Cultural$ and *improvement*. Regarding $Cultural$„ it is shown that its estimator increased by 20% over the precedent model where the instrumental variable was Micro. In the same way, the effect of the number of Transmilenio stations on the level of income of inhabitants increased by 50%.

Finally, the endogeneity test gives as a result the rejection of null hypothesis, which implies that *improvement* is not exogenous and that 2SLS results are more consistent than those found on OLS model. In parallel, the Sargan test and heteroscedasticity tests validate the use of an instrumental variable in the model.

Causality of improvement on the level of income is demonstrated.

4.2.3 Improvement regressed with respect to "Large" enterprises

This final subsection corresponds to results of the three models when the variable *improvement* depends on density, the workforce in each zone and the number of enterprises with more than 50 employees.

As in previous subsections, results for model 1 do not change considerably. The signs, the level of the impact on dependent variable and the statistical significance of estimators remain unchanged. On the side of instrumental variables, the Sargan test and the heteroscedasticity test show that there is no evidence to reject the use of them and that there are no problems of heteroscedasticity. Nonetheless, the endogeneity test determines that there is no causality relationship between the number of Transmilenio stations and the effective size of the labour market. It implies that OLS results are consistent. 2SLS analysis is not verified and the use of instrumental variables is not justified. Endogeneity was not proven..

Table 5: 2SLS results for each interval of time when:
$impro = \pi_0 + \pi_1 WF_i + \pi_2 LE_{si} + u$

	SLM			Income			Employment
	slm 40	slm 50	slm 60	income 40	income 50	income 60	
Intercept	696,094.362***	1012411.954***	1273635.499***	-38,665.391	66,763.779	142,725.122	-17,930.795
	(78,433.247)	(98,263.896)	(99,657.567)	(368,709.311)	(411,342.154)	(478,936.966)	(11,508.759)
Improvment	27,555.484*	45,863.278**	51,261.920**	193,936.451***	198,377.717***	199,520.165***	13,027.699***
	(16,193.431)	(20,287.693)	(20,575.433)	(60,240.124)	(62,034.597)	(63,074.685)	(2,376.114)
SLM				-0.045	-0.134	-0.164	
				(0.376)	(0.316)	(0.320)	
Income	0.043	0.034	0.030				0.003
	(0.034)	(0.043)	(0.043)				(0.005)
Population	-0.420	-0.371	-0.501	-2.944	-2.989	-3.031	0.009
	(0.679)	(0.850)	(0.863)	(2.419)	(2.439)	(2.445)	(0.100)
Area	-0.019*	-0.028**	-0.027**	0.028	0.025	0.025	0.003*
	(0.010)	(0.013)	(0.013)	(0.036)	(0.037)	(0.037)	(0.001)
Mean_dist	-36,037.920***	-46,749.858***	-51,914.820***	-3,937.800	-8,611.248	-10,942.835	-6.675
	(6,685.674)	(8,376.045)	(8,494.842)	(27,409.879)	(28,206.642)	(29,226.119)	(981.010)
Car owners	-2.935	-0.064	1.340	16.169	16.279	16.441	0.821*
	(3.402)	(4.263)	(4.323)	(12.127)	(12.275)	(12.342)	(0.499)
Aver Cost	60.349**	78.544**	89.479***	264.054***	272.547***	276.870***	1.836
	(25.180)	(31.546)	(31.994)	(88.986)	(89.928)	(91.218)	(3.695)
Parks	-295.399	-702.224	-50.733	4,737.804**	4,667.223*	4,751.262**	-33.212
	(680.658)	(852.752)	(864.846)	(2,387.926)	(2,413.226)	(2,415.003)	(99.875)
Social Welfare	-889.780**	-1,055.887**	-908.288*	-2,213.082*	-2,324.718*	-2,339.698*	19.748
	(367.430)	(460.329)	(466.858)	(1,325.650)	(1,334.761)	(1,328.044)	(53.914)
Health	-4,369.952	-6,488.536	-6,201.869	-33,752.496*	-34,428.086*	-34,485.508*	-709.676
	(5,639.089)	(7,064.847)	(7,165.047)	(20,258.235)	(20,469.202)	(20,494.024)	(827.441)
Schools	559.686	479.516	81.976	-10,137.719*	-10,091.636*	-10,102.750*	-395.833
	(1,651.439)	(2,068.981)	(2,098.325)	(5,856.446)	(5,909.067)	(5,930.858)	(242.321)
Cultural	70.010	2,266.227	977.351	28,101.822**	28,410.205**	28,194.966**	1,733.172***
	(3,202.248)	(4,011.887)	(4,068.788)	(11,173.038)	(11,350.086)	(11,307.016)	(469.876)
Recreation	4,843.099	3,502.709	2,667.288	1,346.608	1,701.704	1,801.528	-883.283*
	(3,373.519)	(4,226.461)	(4,286.405)	(12,073.174)	(12,188.071)	(12,222.204)	(495.007)
Neighborhoods	-704.184	-847.041	-1,215.219**	183.564	103.703	18.774	-57.692
	(487.194)	(610.374)	(619.031)	(1,779.510)	(1,795.660)	(1,822.401)	(71.488)
Endogeneity - Wu - Hausman F Test	0.66310	1.91267	2.60483	15.17193 ***	14.91195***	14.56189 ***	417.58137***
Sargan	3.166	3.391	1.342	0.985	1.001	1.104	1.553
Heterocedasticity	19.129	16.858	13.779	16.528	16.328	16.424	7.043
R^2	0.568	0.561	0.566	0.101	0.085	0.081	-0.725

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Author calculations

Concerning the results for model 2, table 5 reveals the same conclusions as when a variable suspected to be endogenous is regressed with respect to the number of middle-size enterprises. This denotes that the variable *Health* has the bigger negative impact on inhabitants' level of income. As explained above,

it could imply that inhabitants with higher incomes chose to live in zones with few establishments that provide health care. The average income of people living in areas with one fewer entity that provides health services with respect to the average income of people living in areas with an additional facility of this kind is 34,500 COP higher (8% of the minimum wage).

Concerning the impact of the number of Transmilenio stations on the income of inhabitants, results reveal that first of all, this variable is by far that which considerably impacts the income of inhabitants. As in a previous subsection, it signifies that the income of inhabitants with an additional station in a zone will be on average 197.000 COP (95 USD) higher.

In effect, concerning the endogeneity test, table 5 reveals that the number of TM stations is endogenous. It suggests that a causal relationship of the variable *improvement* and the level of income is verified when the endogenous variable depends on the number of large enterprises in each zone.

In the same way, results for model 3 suggest the same interpretations. In fact, the variable *employment* also depends considerably on the variable *improvement*. An additional TM in a zone will boost the number of jobs in this zone by 13.027. As in model 2, the level of enhancement of public transport is endogenous, which means there is a causal relationship between the number of stations of TM on the number of jobs on each zone. For both models, there is no evidence against the use of the instruments and there are no problems of heteroscedasticity.

Causal relations are confirmed in models 2 and 3 when the number of "large" enterprises plays the role of instrumental variable.

Table 6: Summary of causality test results

	Test	SLM	Income	Employment
Micro	Endogeneity	No	No	Yes
	Sargan	No	No	No
	Heteroscedasticity	No	Yes	No
PME	Endogeneity	No	Yes	Yes
	Sargan	No	No	No
	Heteroscedasticity	No	No	No
Large	Endogeneity	No	Yes	Yes
	Sargan	No	No	No
	Heteroscedasticity	No	No	No

Author calculations

5 CONCLUSION

From the beginning the study has focused on the fact that public transport improvements affect job accessibility, inhabitants' income and the number of jobs in each zone of the city. In the same way, it was suggested that enhancements of public transports could have a causal relation with income, employment and job accessibility, and that was the reason to analyse the possible endogeneity of this variable. The proxy for improvement of public transport was defined by the number of stations of TM in each zone. This variable, which is suspected to be the endogenous one, is defined by some instruments on each model that their use appear to be accepted by different test (Sargan test Heteroscedasticy test).

Despite the lack of data for several periods, this "cross section" analysis displays interesting results that suggest some particular conclusions that have to be enounced.

Compilation of results reveals that on OLS regressions for the effective size of labour market, improvement to public transport always has a direct and important effect on dependent variables with a statistically significant level. This corroborates findings of previous studies (Ihlanfeldt and Sjoquist, 1991).

In fact, this variable was always the one that had the biggest positive impact on the three different models, regardless of the commuting time, but its amplitude appears to be lower than expected and in some cases its statistical significance may not reveal its real importance (model 2).

Those results motivate the use of the 2SLS method, which explains the suggestion of instrumental variables in order to capture the real and thus suspected impact of the number of stations of Transmilenio on dependent variables.

Regarding the size of the labour market, variables that had a bigger but a negative effect on the $ESLM$ were the mean distance travelled and the area of each zone, consistent with Prud'Homme and Lee (1999) with respect to the "three S" (Speed, Sprawl and Spread). Endogeneity of improvement of public transport on the $ESLM$ was not proven at all. In effect, regardless of the size of companies in each zone and the workforce in each zone (instrumental variables), the number of TM stations in each zone has no causal effect on the number of jobs reachable within 40, 50 and 60 minutes. This may imply that, for the case of Bogotá, the decision to enhance public transport in some zones of the city – or, to be more precise, the decision to construct a TM station in a zone – could not be motivated by the jobs potentially reachable from these zones. The $ESLM$ depends "significantly" on the number of Transmilenio stations but there is no evidence to assert that the degree of enhancement of public transport has a causality effect on the size of the labour market for inhabitants.

In the other hand, as OLS results for model 2 reveal, the number of stations of Transmilenio in each zone of the city has no statistical influence on income.

Nevertheless, this study supposes that this variable should have a greater statistical influence on income. In effect, after applying 2SLS method and excepting when the endogenous variable is defined by the number of micro enterprises (instrumental variable), results show that improvement to public transport is significant and has an important impact on the income of inhabitants.

In addition, when instrumental variables are the workforce in each zone and the number of small and middle-size enterprises or the number of large enterprises, enhancements of public transport are endogenous to the income; thus a causality relation of improvements of public transport to income is demonstrated. Furthermore, the use of an instrumental variable reveals an underestimation of the effect of the number of Transmilenio stations on the income of inhabitants.

These results allow some suggestions. First of all it might be said that people with the highest incomes are those that can settle in zones with a high number of Transmilenio stations. Second, in order to understand the results, microenterprises appear to be established in low income zones, contrary to small, medium and large enterprises that seem to settle in zones with high levels of income. Those conclusions suggest that small, medium and large enterprises seem to be in a bigger capacity, like people with high incomes, to settle on zones benefiting more from enhancements of public transports. Until 2007, zones that have more Transmilenio stations appeared to attract people and enterprises with the highest incomes in the city.

In relation to employment, this study showed that improvement to public transport is always statistically significant. Moreover, its influence increases radically when the number of jobs in each zone depends on instrumental variables. Likewise, this is even more evident if the size of enterprises increases. As with income, the number of Transmilenio stations is endogenous to the number of jobs; hence its causality relationship is verified no matter the size of the companies. In a first time, results suggest that public transports improvements incentive the creation of jobs in a zone. Nevertheless, it could also suggest that jobs (and not workers) decide to settle on zones with better level of public transport improvements. As denoted previously, the biggest enterprises, which are those that create almost 70% of employment in the city, prefer to be close to Transmilenio stations. Microenterprises seem not to be capable of benefiting from public transport enhancements.

The results shed light on the causal effect of improvement to the public transport system on the income and number of jobs in each zone of the city of Bogotá. In particular, they indicate that the greater the presence of Transmilenio stations, the bigger will be the income and the number of jobs. Moreover, this causal relationship is even greater if the size of companies that settle in those zones is bigger. For now, zones where the largest enterprises are located benefit from urban transportation enhancements. After testing this causality, this analysis might suggest to build and expand the Transmilenio system to zones where microenterprises, which assure 30% of formal jobs in the city, are established.

6 Annexes

Annex 1 Average commuting time according to type of transport

Conveyance	Commuting time (min)
By walk	27
Bicycle	31
Other	31
Taxi	34
Moto	35
Private cars	40
Transmilenio	41
Truck	46
Intermunicipal bus	46
Scolar bus	48
Private bus	49
Public transport diferent of TM	56
Average	40

Author calculations

Annex 2: Correlation Matrix

		1 ESLM_60	2 Income	3 Employment	4 Improvment	5 Population	6 Area	7 Mean_dist
1	ESLM_60	1,00						
2	Income	0,31	1,00					
3	Employment	0,36	0,44	1,00				
4	Improvment	0,38	0,20	0,56	1,00			
5	Population	-0,24	-0,28	-0,03	0,09	1,00		
6	Area	-0,43	-0,01	0,05	-0,01	0,48	1,00	
7	Mean_dist	-0,41	0,13	-0,07	-0,16	-0,34	0,19	1,00
8	Car owners	0,04	0,00	0,06	0,14	0,57	0,24	-0,27
9	Aver Cost	0,21	0,47	0,17	0,05	-0,31	0,04	0,35
10	Parks	-0,21	-0,05	-0,03	0,06	0,74	0,43	-0,25
11	Social Welfare	-0,30	-0,38	-0,03	-0,01	0,65	0,21	-0,26
12	Health	0,00	-0,17	0,20	0,27	0,45	0,17	-0,29
13	Schools	0,08	-0,03	0,33	0,36	0,70	0,27	-0,32
14	Cultural	-0,11	-0,07	0,34	0,05	0,45	0,12	-0,21
15	Recreation	0,20	0,08	0,32	0,46	0,56	0,21	-0,32
16	Neighborhoods	-0,32	-0,02	-0,03	0,02	0,60	0,53	-0,06

	8 Car owners	9 Aver Cost	10 Parks	11 Social Welfare	12 Health	13 Schools	14 Cultural	15 Recreation	16 Neighborhoods
1									
2									
3									
4									
5									
6									
7									
8	1,00								
9	-0,02	1,00							
10	0,45	-0,23	1,00						
11	0,24	-0,36	0,49	1,00					
12	0,22	-0,22	0,33	0,40	1,00				
13	0,55	-0,11	0,53	0,46	0,53	1,00			
14	0,13	-0,20	0,39	0,54	0,38	0,64	1,00		
15	0,51	-0,06	0,47	0,43	0,49	0,81	0,38	1,00	
16	0,37	-0,06	0,63	0,30	0,22	0,42	0,32	0,30	1,00

Author calculations

Annex 3: VIF calculations

SLM		
Variable	VIF	1/VIF
education	7.72	0.129463
pop_2007	6.11	0.163585
rec_sport	4.27	0.234157
culture	3.64	0.274602
parks	2.95	0.338479
soc_welf	2.75	0.364112
employment	2.56	0.390699
number_neigh	2.15	0.465232
Improvment	2.07	0.482405
income	2.02	0.494263
have_a_car	2.02	0.494576
superficie	1.94	0.516430
health	1.58	0.631896
aver_cost_~1	1.58	0.632610
mean_dist	1.55	0.645666
Mean VIF	3.00	

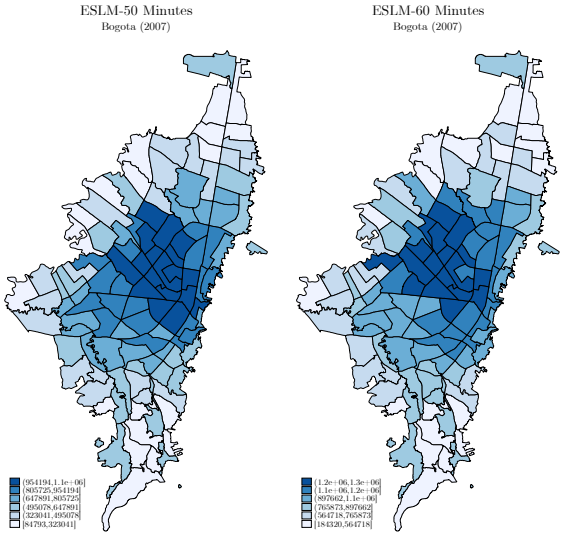
Income		
Variable	VIF	1/VIF
education	7.71	0.129664
pop_2007	5.98	0.167186
rec_sport	4.23	0.236516
culture	3.69	0.270689
parks	2.76	0.361931
soc_welf	2.74	0.365410
slm_60_pt	2.70	0.370087
employment	2.31	0.432448
number_neigh	2.18	0.457705
superficie	2.12	0.470744
mean_dist	2.08	0.481033
Improvment	2.08	0.481326
have_a_car	2.02	0.494343
health	1.56	0.640362
aver_cost_~1	1.56	0.642255
Mean VIF	3.05	

Employment		
Variable	VIF	1/VIF
education	7.72	0.129567
pop_2007	6.08	0.164413
rec_sport	4.27	0.234169
culture	3.04	0.328818
parks	2.86	0.349210
soc_welf	2.74	0.364398
number_neig	2.11	0.474836
have_a_car	2.00	0.498964
superficie	1.82	0.547994
income	1.73	0.577508
aver_cost_~1	1.58	0.632700
health	1.58	0.633736
mean_dist	1.54	0.649947
Improvment	1.50	0.666793
Mean VIF	2.90	

Author calculations

Annex 4: Size of Labor Market respecting the zone where people live (UPZ)

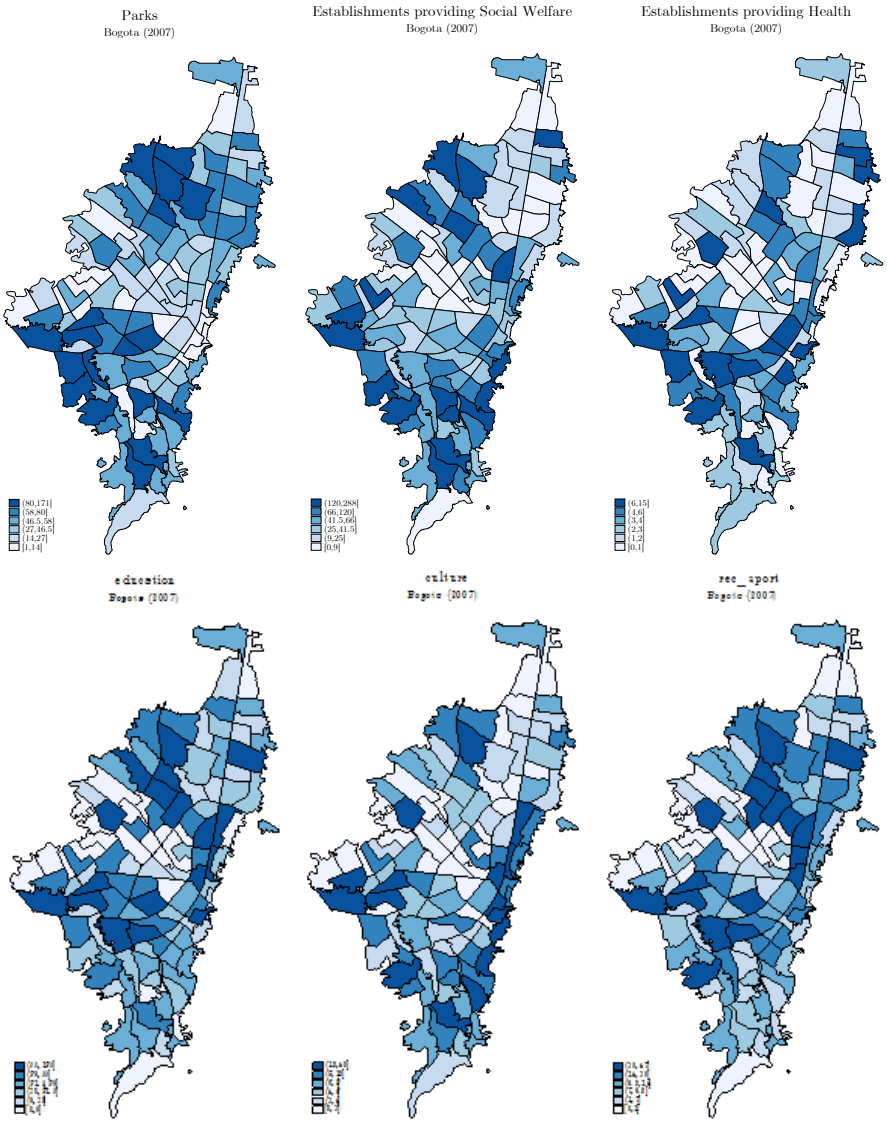
Bogota (2007)



Author calculations

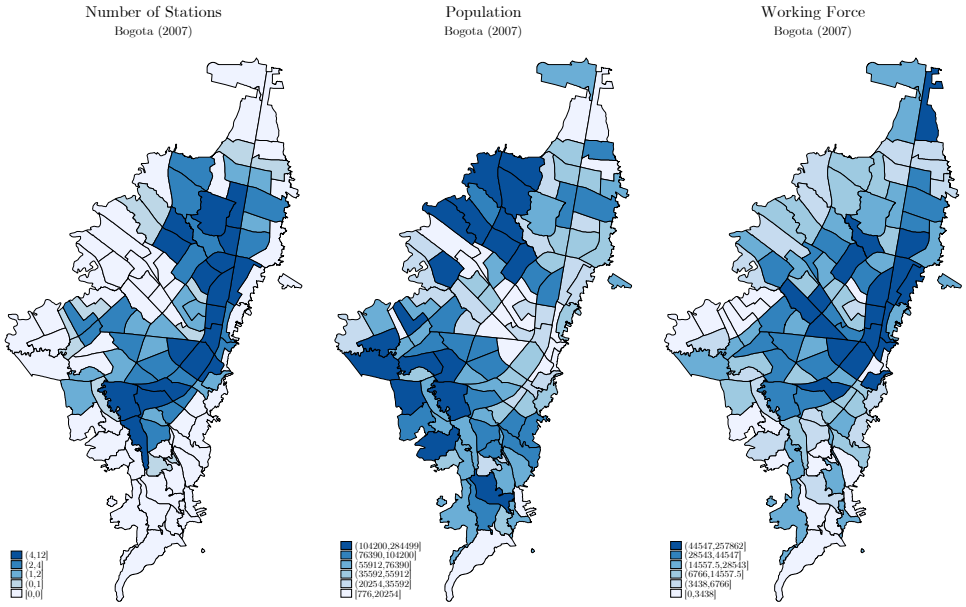
Annex 5: Amenities by Zones (UPZ)

Bogota (2007)



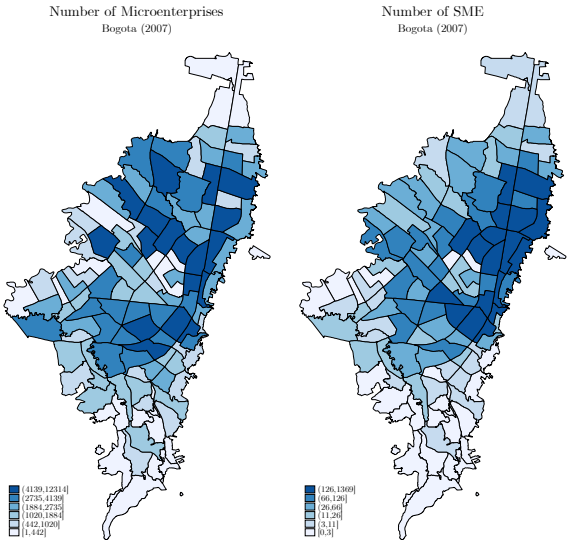
Author calculations

Table 6: Concentration of Stations of TM, Population, Workforce configuration by (UPZ)
Bogotá (2007)



Author calculations

Annex 7: Concentration of enterprises by (UPZ)
Bogota (2007)



Author calculations

7 Bibliography

Attanasi, Giuseppe, Casoria, Fortuna, Centorrino, Samuele and Urso, Giulia, (2012), Cultural investment, local development and instantaneous social capital: A case study of a gathering festival in the South of Italy, No 12-291, TSE Working Papers, Toulouse School of Economics (TSE), http://EconPapers.repec.org/RePEc:tse:wpaper:25758.

Bagley Michael N.. Mokhtarian, Patricia L. (2002). The impact of neighborhood type on travel behavior: A structural equations modeling approach. The annals of regional Science, 36(5), 279-297.

Bocarejo, J. y Oviedo, D. (2010), "Transport Accessibility and Social Exclusion: A Better Way to Evaluate Public Transport Investment?" [en línea], disponible en http://www.ucl.ac.uk/dpu/metrocables/ dissemination / Bocarejo_y_Oviedo.pdf, consultado el 27 de septiembre de 2011.

Cao, Xinyu; Mokhtarian, Patricia L.; & Handy, Susan L.(2007). Do changes in neighborhood characteristics lead to changes in travel behavior ? A structural equations modeling approach. Transportation, 34(5), 535 - 556. Retrieved from: http://escholarship.org/uc/item/49393360

Cervero R, Sandoval O, Landis J (2000) Transportation as a stimulus to welfare-to-work: private versus public mobility. J Plan Edu Res 22:50–63

David, Quentin, Janiak, Alexandre and Wasmer, Etienne, (2010), Local social capital and geographical mobility, Journal of Urban Economics, 68, issue 2, p. 191-204.

Hansen, W. G. (1959), "How Accessibility Shapes Land Use", en Journal of American Institute of Planners, No. 25, pp. 73-76.

Ihlanfeldt, K., and D. Sjoquist. 1991. The role of space in determining the occupation of black and white workers. Regional Science and Urban Economics 21:295-99.

Ihlanfeldt, Keith R., Sjoquist, David L. 1998 : "The spatial mismatch hypothesis: A review of recent studies and their implications for welfare reform". Housing Policy Debate , Vol. 9, Iss. 4.

Kawabata, M. (2003) "Job access and employment among low-skilled autoless workers in U.S. Metropolitan areas." Environment and Planning A, Vol. 35, No. 9, pp. 1651-1668.

Krugman, P. (1997), Development, Geography and Economic Theory, Cambridge, The MIT Press.

Koenig, G. (1974), "Théorie économique de l'accessibilité urbaine", en Revue économique, Vol. 25, No. 2, pp. 275-279.

Prud'homme, R. y Lee, C. W. (1999), "Size, Sprawl, Speed and the Efficiency of Cities", en Urban Studies, Vol. 36, No.11, pp. 1849-1858.

Shengyi Gao, Patricia L. Mokhtarian, Robert A. Johnston (Exploring connection among job access, emply, income and auto ownership using SEM, 2007)

Wenglenski, S. (2004), "Une mesure des disparités sociales d'accessibilité à l'emploi", en Revue l'Economie Régionale et Urbaine (RERU), No. 4, pp. 153-169.